INTERNET-DRAFT   D. Meyer
draft-ietf-grow-collection-communities-04.txt
Category   Best Current Practice
Expires: September 2004   March 2004

BGP Communities for Data Collection
<draft-ietf-grow-collection-communities-04.txt>

Status of this Memo

This document is an Internet-Draft and is subject to all provisions of Section 10 of RFC2026.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

The list of current Internet-Drafts can be accessed at http://www.ietf.org/1id-abstracts.html

The list of Internet-Draft Shadow Directories can be accessed at http://www.ietf.org/shadow.html

This document is a product of the GROW WG. Comments should be addressed to the authors, or the mailing list at grow@lists.uoregon.edu.

Copyright Notice

Abstract

BGP communities (RFC 1997) are used by service providers for many purposes, including tagging of customer, peer, and geographically originated routes. Such tagging is typically used to control the scope of redistribution of routes within a provider's network, and to its peers and customers. With the advent of large scale BGP data collection (and associated research), it has become clear that the information carried in such communities is essential for a deeper understanding of the global routing system. This document defines standard (outbound) communities and their encodings for export to BGP route collectors.

Table of Contents

1. Introduction

BGP communities [RFC1997] are used by service providers for many purposes, including tagging of customer, peer, and geographically originated routes. Such tagging is typically used to control the scope of redistribution of routes within a providers network, and to its customers and peers. Communities are also used for a wide variety of other applications, such as allowing customers to set attributes such as LOCAL_PREF [RFC1771] by sending appropriate communities to their service provider. Other applications include signaling various types of VPNs (e.g., VPLS [VPLS]), and carrying link bandwidth for traffic engineering applications [EXTCOMM].

With the advent of large scale BGP data collection [RIS,ROUTEVIEWS] (and associated research), it has become clear that the geographical and topological information, as well as the relationship the provider has to the source of a route (e.g., transit, peer, or customer), carried in such communities is essential for a deeper understanding of the global routing system. This document defines standard communities for export to BGP route collectors. These communities represent a significant part of information carried by service providers as of this writing, and as such could be useful for internal use by service providers. However, such use is beyond the scope of this memo. Finally, those involved in BGP data analysis are encouraged to verify with their data sources as to which peers implement this scheme (as there is a large amount of existing data as well as many legacy peerings).

The remainder of this document is organized as follows. Section 2 provides both the definition of terms used as well as the semantics of the communities used for BGP data collection, and section 3 defines the corresponding encodings for RFC 1997 [RFC1997] communities. Finally, section 4 defines the encodings for use with extended communities [EXTCOMM].

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14, RFC 2119 [RFC2119].

2. Definitions

In this section, we define the terms used and the categories of routes that may be tagged with communities. This tagging is often

referred to coloring, and we refer to a route's "color" as its community value. The categories defined here are loosely modeled on those described in [WANG] and [HUSTON].

2.1. Peers and Peering

Consider two network service providers, A and B. Service providers A and B are defined to be peers when (i). A and B exchange routes via BGP, and (ii). traffic exchange between A and B is settlement-free. This arrangement is also typically known as "peering". Peers typically exchange only their respective customer routes (see "Customer Routes" below), and hence exchange only their respective customer traffic. See [HUSTON] for a more in-depth discussion of the business models surrounding peers and peering.

2.2. Customer Routes

Customer routes are those routes which are heard from a customer via BGP and are propagated to peers and other customers. Note that a customer can be an enterprise or another network service provider. These routes are sometimes called client routes [HUSTON].

2.3. Peer Routes

Peer routes are those routes heard from peers via BGP, and not propagated to other peers. In particular, these routes are only propagated to the service provider's customers.

2.4. Internal Routes

Internal routes are those routes that a service provider originates and passes to its peers and customers. These routes are frequently taken out of the address space allocated to a provider.

D. Meyer. Section 2.4. [Page 5]

2.5. Internal More Specific Routes

Internal more specific routes are those routes which are frequently used for circuit load balancing purposes, IGP route reduction, and also may correspond to customer services which are not visible outside the service provider's network. Internal more specific routes are not exported to any external peer.

2.6. Special Purpose Routes

Special purpose routes are those routes which do not fall into any of the other classes described here. In those cases in which such routes need to be distinguished, a service provider may color such routes with a unique value. Examples of special purpose routes include anycast routes, and routes for overlay networks.

2.7. Upstream Routes

Upstream routes are typically learned from upstream service provider as part of a transit service contract executed with the upstream provider.

2.8. National Routes

These are route sets that are sourced from and/or received within a particular country.

2.9. Regional Routes

Several global backbones implement regional policy based on their deployed footprint, and on strategic and business imperatives. Service providers often have settlement free interconnections with an AS in one region, and that same AS is a customer in another region. This mandates use of regional routing, including community attributes

D. Meyer. Section 2.9.

set by the network in question to allow easy discrimination among regional routes. For example, service providers may treat a route set received from another service provider in Europe differently than the same route set received in North America, as it is common practice to sell transit in one region while peering in the other.

3. RFC 1997 Community Encoding and Values

In this section we provide RFC 1997 [RFC1997] community values for the categories described above. RFC 1997 communities encoded as BGP Type Code 8, and are treated as 32 bit values ranging from 0x0000000 through 0xFFFFFFF. The values 0x0000000 through 0x0000FFFF and 0xFFFF0000 through 0xFFFFFFFF are reserved.

The best current practice among service providers is to use the high order two octets to represent the providers AS number, and the low order two octets to represent the classification of the route, as depicted below:

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|            <AS>               |          <Value>              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

where <AS> 16 bit AS number, and <Value> is the encoding of the value. For example, the encoding 0x2A7C029A would represent the AS 10876 with value 666.

3.1. Community Values for BGP Data Collection

In this section we define the RFC 1997 community encoding for the route types described above for use in BGP data collection. It is anticipated that a service provider's internal community values will be converted to these standard values for output to a route collector.

This document follows the best current practice of using the basic format <AS>:<Value>. The values for the route categories are described in the following table:

D. Meyer. Section 3.1.

Category	Value
Reserved	<AS>:0000000000000000
Customer Routes	<AS>:0000000000000001
Peer Routes	<AS>:0000000000000010
Internal Routes	<AS>:0000000000000011
Internal More Specific Routes	<AS>:0000000000000100
Special Purpose Routes	<AS>:0000000000000101
Upstream Routes	<AS>:0000000000000110
Reserved	<AS>:0000000000000011- <AS>:0000111111111111
National and Regional Routes	<AS>:0001000000000000- <AS>:1111111111111111
Africa (AF)	<AS>:0001<X><CC>
Oceania (OC)	<AS>:0010<X><CC>
Asia (AS)	<AS>:0011<X><CC>
Antarctica (AQ)	<AS>:0100<X><CC>
Europe (EU)	<AS>:0101<X><CC>
Latin America/Caribbean islands (LAC)	<AS>:0110<X><CC>
North America (NA)	<AS>:0111<X><CC>
Reserved	<AS>:1000000000000000- <AS>:1111111111111111

In the above table,

<AS> is the 16-bit AS
<R> is the 5-bit Region
<X> is 1-bit satellite link indication (1 if satellite link, 0 otherwise)
<CC> is the 10-bit ISO-3166-2 country code

That is:

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|             <AS>              |   <R>   |X|        <CC>       |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

For example, the encoding for a national route over a terrestrial link in AS 10876 from the Fiji Islands would be:

<AS> = 10876 = 0x2A7B
<R> = OC = 0010
<X> = 0x0
<CC> = Fiji Islands Country Code = 242 = 0011110010

so that the low order 16 bits look like 001000011110010 = 0x10F2.

D. Meyer. Section 3.1. [Page 8]

```
    0                   1                   2                   3
    0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |           0x2A7C              |           0x10F2              |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Note that a configuration language might allow the specification of this community as 10876:4338 (0x1F2 == 4338 decimal).

Finally, note that these categories are not intended to be mutually exclusive, and multiple communities can be attached where appropriate.

4. Extended Communities

In some cases, the encoding described in section 3.1 may clash with a service provider's existing community assignments. Extended communities [EXTCOMM] provide a convenient mechanism that can be used to avoid such clashes.

The Extended Communities Attribute is a transitive optional BGP attribute with the Type Code 16, and consists of a set of extended communities of the following format:

```
    0                   1                   2                   3
    0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |  Type high    |  Type low(*)  |                               |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+          Value                |
   |                                                               |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

For purposes of BGP data collection, we encode the communities described in section 3.1 using the two-octet AS specific extended community type, which has the following format:

```
    0                   1                   2                   3
    0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |      0x00     |   Sub-Type    |    Global Administrator       |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                     Local Administrator                       |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

The two-octet AS specific extended community attribute encodes the

service provider's two octet Autonomous System number (as assigned by an Regional Internet Registry, or RIR) in the Global Administrator field, and the Local Administrator field may encode any information.

This document assigns Sub-Type 0x05 for BGP data collection, and specifies that the <Value> field, as defined in section 3.1, is carried in the low order octets of the Local Administrator field. The two high order octets of the Local Administrator field are reserved, and are set to 0x00 when sending and ignored upon receipt.

For example, the extended community encoding for 10876:4338 (representing a terrestrial national route in AS 10876 from the Fiji Islands) would be:

```
   0                   1                   2                   3
   0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
  +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
  |      0x00     |      0x05     |           0x2A7C              |
  +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
  |      0x00     |      0x00     |           0x10F2              |
  +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

4.1. Four-octet AS specific extended communities

The four-octet AS specific extended community is encoded as follows:

```
   0                   1                   2                   3
   0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
  +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
  |      0x02     |    0x05       |    Global Administrator       |
  +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
  | Global Administrator (cont.)  |           0x10F2              |
  +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

In this case, the 4 octet Global Administrator sub-field contains a 4-octets Autonomous System number assigned by the IANA.

5. Acknowledgments

The community encoding described in this document germinated from an interesting suggestion from Akira Kato at WIDE. In particular, the idea would be to use the collection community values to select paths

that would result in (hopefully) more efficient access to various services. For example, in the case of RFC 3258 [RFC3258] based DNS anycast service, BGP routers may see multiple paths to the same prefix, and others might be coming from the same origin with different paths, but others might be from different region/country (with the same origin AS).

Joe Abley, Randy Bush, Sean Donelan, Xenofontas Dimitropoulos, Vijay Gill, John Heasley, Geoff Huston, Steve Huter, Olivier Marce, Ryan McDowell, Rob Rockell, Rob Thomas, Pekka Savola, and Patrick Verkaik all made many insightful comments on early versions of this draft. Henk Uijterwaal suggested the use of the ISO-3166-2 country codes.

D. Meyer. Section 5.

6. Security Considerations

While this document introduces no additional security considerations into the BGP protocol, the information contained in the communities defined in this document may in some cases reveal network structure that was not previously visible outside the provider's network. As a result, care should be taken when exporting such communities to route collectors. Finally, routes exported to a route collector should also be tagged with the NO_EXPORT community (0xFFFFFF01).

6.1. Total Path Attribute Length

The communities described in this document are intended for use on egress to a route collector. Hence an operator may choose to overwrite its internal communities with the values specified in this document when exporting routes to a route collector. However, operators should in general ensure that the behavior of their BGP implementation is well-defined when the addition of an attribute causes a PDU to exceed 4096 octets. For example, since it is common practice to use community attributes to implement policy (among other functionality such as allowing customers to set attributes such as LOCAL_PREF), the behavior of an implementation when the attribute space overflows is crucial. Among other behaviors, an implementation might usurp the intended attribute data or otherwise cause indeterminate failures. These behaviors can result in unanticipated community attribute sets, and hence result in unintended policy implications.

7. IANA Considerations

This document assigns a new Sub-Type for the AS specific extended community type. In particular, the IANA should assign Sub-type 0x05, using the "First Come First Served" policy defined in RFC 2434 [RFC2434], for the Sub-Type defined in Section 4. This corresponds to a Type Field value of 0x0005.

8. References

8.1. Normative References

[EXTCOMM] Sangali, S., D. Tappan and Y. Rekhter, "BGP Extended Communities Attribute", draft-ietf-idr-bgp-ext-communities-06.txt, Work in progress.

[ISO-3166-2] http://www.iso.org/iso/en/prods-services/iso3166ma/index.html

[RIS-ISO-3166] ftp://ftp.ripe.net/iso3166-countrycodes.txt

[RFC1771] Rekhter, Y. and T. Li (Editors), "A Border Gateway Protocol (BGP-4)", RFC 1771, March 1995.

[RFC1997] Chandra, R. and P. Traina, "BGP Communities Attribute", RFC 1997, August 1996.

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

8.2. Informative References

[HUSTON] Huston, G., "Interconnection, Peering, and Settlements", http://www.isoc.org/inet99/proceedings/1e/1e_1.htm

[RFC2028] Hovey, R. and S. Bradner, "The Organizations Involved in the IETF Standards Process", BCP 11, RFC 2028, October 1996.

[RFC2434] Narten, T., and H. Alvestrand, "Guidelines for Writing an IANA Considerations Section in RFCs", BCP 26, RFC 2434, October 1998.

[RFC3258] Hardie, T., "Distributing Authoritative Name Servers via Shared Unicast Addresses", RFC 3258, April 2002.

[RIS] "Routing Information Service", http://www.ripe.net/ris

[ROUTEVIEWS] "The Routeviews Project", http://www.routeviews.org

D. Meyer. Section 8.2. [Page 13]

[VPLS] Kompella, K., et al., "Virtual Private LAN Service", draft-ietf-l2vpn-vpls-bgp-01.txt, Work in Progress.

[WANG] Wang, F. and L. Gao, "Inferring and Characterizing Internet Routing Policies", ACM SIGCOMM Internet Measurement Conference 2003.

9. Author's Addresses

David Meyer
EMail: dmm@1-4-5.net

10. Full Copyright Statement

11. Intellectual Property

The IETF takes no position regarding the validity or scope of any Intellectual Property Rights or other rights that might be claimed to pertain to the implementation or use of the technology described in this document or the extent to which any license under such rights might or might not be available; nor does it represent that it has made any independent effort to identify any such rights. Information on the procedures with respect to rights in RFC documents can be found in BCP 78 and BCP 79.

Copies of IPR disclosures made to the IETF Secretariat and any assurances of licenses to be made available, or the result of an attempt made to obtain a general license or permission for the use of such proprietary rights by implementers or users of this specification can be obtained from the IETF on-line IPR repository at http://www.ietf.org/ipr.

The IETF invites any interested party to bring to its attention any copyrights, patents or patent applications, or other proprietary rights that may cover technology that may be required to implement this standard. Please address the information to the IETF at ietf-ipr@ietf.org.

12. Acknowledgement

Funding for the RFC Editor function is currently provided by the Internet Society.